a sampling device that stores summarized traffic data that describe each occurrence of the digital content in the traffic data; and

an accessing device that presents the traffic data and the summarized traffic data to a user.

2. (Amended One Time) The system of claim 1, wherein the estimating device retrieves the traffic data from at least one proxy cache server.

A1

3. (Amended One Time) The system of claim 1, wherein the sampling device computes the number of impressions of the digital content for a web site on the network.

4. (Amended One Time) The system of claim 1, wherein the sampling device includes:

a prober that fetches a web page from the network;

an extractor that locates a fragment of the web page that includes the digital content; and

a classifier that performs a structural analysis of the fragment to classify the digital content.

5. (Amended One Time) The system of claim 1, wherein the accessing device generates a report when the traffic data or the summarized traffic data satisfy at least one criterion.

6. (Amended One Time) A method of estimating prevalence of digital content on a network, comprising the steps of:

estimating the global traffic to a at least one Web site on the network to provide traffic data;

statistically sampling the contents of said said at least one Web site to provide sampling data;

storing the traffic data and the sampling data;

accessing the traffic data and the sampling data to generate a report.

7. (Amended One Time) A system for estimating the prevalence of digital content on a network connected to at least one network site that includes at least one network server to access at least one uniform resource locator, the system comprising:

a database;

a traffic analysis system that stores traffic data from a traffic sampling system in the database, the traffic data including said at least one uniform resource locator;

a digital content sampling system that stores the digital content at said at least one uniform resource locator in the database; and

a statistical summarization system that stores summarization data that describe the digital content in the database.

8. (Amended One Time) The system of claim 7, further comprising:

a Web front end that connects to the network and the database, wherein a client client uses a browser to connect to the Web front end.

9. (Amended One Time) The system of claim 7, further comprising:

a user interface that an account manager, an operator, or a media editor can use to administer the system.

10. The system of claim 7, wherein the network is the Internet, and wherein the network site is a A1 Web site.

11. The system of claim 7, wherein the traffic analysis system further comprises:

an anonymity system that receives the traffic data sample from the traffic sampling system and produces a clean traffic data sample; and

a traffic summarization system that produces a summarization of the clean traffic data sample and stores the traffic data sample in the database.

12. (Amended One Time) The system of claim 11, wherein the anonymity system produces a clean traffic data sample by removing a network address or a cookie from the traffic data sample.

13. The system of claim 11, wherein the summarization of the clean traffic data sample includes a reference to said at least one uniform resource locator and a tally of the number of times said at least one uniform resource locator was requested.

14. (Amended One Time) The system of claim 7, wherein the digital content sampling system further comprises:

a probe mapping system that uses the summarization data to create a probe map for the network, the probe map including a mapping for said at least one uniform resource locator;

a uniform resource locator retrieval system that retrieves said at least one uniform resource locator from the network server;

A1 a browser emulation environment that conducts a simulation of the display of said at least one uniform resource locator in a browser;

a digital content extractor that stores the digital content from said at least one uniform resource locator in the database; and

a structural classifier that stores at least one classification type for the digital content in the database.

15. (Amended One Time) The system of claim 14, wherein the probe map further comprises:

a probability of the likelihood that said at least one uniform resource location will be sampled; and

a scale that determines the contribution of said at least one uniform resource location to the summarization data.

16. The system of claim 14, wherein the simulation includes executing a program embedded in said at least one uniform resource locator.

17. (Amended One Time) The system of claim 16, wherein the program is a JavaScript script, a Java applet, a Perl script, or a common gateway interface program.

A1 18. The system of claim 14, wherein the simulation includes executing dynamic digital content in said at least one uniform resource locator.

19. (Amended One Time) The system of claim 18, wherein the dynamic content is an interlaced GIF image, an MPEG movie, or an MP3 audio file.

20. (Amended One Time) The system of claim 14, wherein the digital content extractor retrieves the digital content from said at least one uniform resource locator by applying a rule set defined by a media editor.

21. (Amended One Time) The system of claim 14, wherein the digital content extractor retrieves the digital content from said at least one uniform resource locator by using an automated digital content detection system.

22. (Amended One Time) The system of claim 21, wherein the automatic digital detection system comprises:

a structural detector that locates an XML structure; and

a feature detector that locates an XML feature within the XML structure.

23. (Amended One Time) The system of claim 14, wherein the structural classifier determines said at least one classification type for the digital content.

24. (Amended One Time) The system of claim 7, wherein the user interface further comprises:

a system account management interface that assists the account manager with creating and modifying an account on the system;

a site administration interface that assists the operator with the administration of said at least one network site;

a taxonomy administration interface that assists the media editor with the administration of the taxonomy data;

a digital content classification interface that assists the media editor with the classification of the digital content; and

a rate card collection interface that assists the media editor with the administration of the rate card data.

25. (Amended One Time) A system for estimating prevalence of digital content on a network, comprising:

a memory device; and

a processor disposed in communication with the memory device, the processor configured to:

obtain traffic data from at least one Web site on the network;

compute a number of impressions for the digital content in the traffic data;

retrieve the digital content from the traffic data to generate sampling data; and

generate prevalence estimates for the digital content from the traffic data and the sampling data.

26. (Amended One Time) The system of claim 25, wherein the processor is further configured to:

retrieve a Web page from said at least one Web site:

extract a fragment from the Web page; and

classify the fragment.

27. (Amended One Time) The system of claim 25, wherein the processor is further configured to:

generate the traffic data by retrieving anonymous traffic data.

28. (Amended One Time) The system of claim 27, wherein the processor is further configured to:

retrieve anonymous data samples by removing data from the traffic data that identifies a user on the network.

A1

29. (Amended One Time) The system of claim 25, wherein the processor is further configured to:

classify a fragment within the sampling data.

30. (Amended One Time) The system of claim 29, wherein the processor is further configured to:

classify the fragment by analyzing the fragment for uniqueness and adding information to a database regarding the uniqueness of the fragment.

31. (Amended One Time) The system of claim 30, wherein the processor is configured to:

classify the fragment by detecting a duplicate fragment.

32. (Amended One Time) The system of claim 25, wherein the processor is further configured to:

interact with a user interface that administers the system.

33. (Amended One Time) The system of claim 25, wherein the processor is further configured to:

include uniform resource locator information regarding said at least one Web site in the traffic data.

A1

34. (Amended One Time) The system of claim 25, wherein the processor is further configured to:

perform data integrity monitoring of the sampling data.

35. (Amended One Time) The system of claim 25, wherein the processor is further configured to:

serve as an automatic digital content detection system.

36. (Amended One Time) The system of claim 35, wherein the automatic advertisement detection system applies at least one heuristic algorithm to detect digital content within an HTML or an XML document and normalizes the detected HTML or XML content into a hierarchical form.

37. (Amended One Time) A method for using a computer to estimate prevalence of digital content on a network, comprising the steps of:

obtaining traffic data from at least one Web site on the network;

computing a number of impressions for the digital content from said at least one Web site;

retrieving the digital content from the traffic data to generate sampling data; and

generating prevalence estimates for the digital content from the traffic data and the sampling data.

A1

38. (Amended One Time) The method of claim 37, wherein retrieving the digital content further comprises the steps of:

retrieving a Web page from said at least one Web site;

extracting a fragment from the Web page; and

classifying the fragment.

39. (Amended One Time) The method of claim 37, wherein the traffic data is anonymous.

40. (Amended One Time) The method of claim 39, wherein the traffic data is made anonymous by removing data from the traffic data that identifies a user on the network.

41. (Amended One Time) The method of claim 37, further comprising the step of:

classifying a fragment within the sampling data.

42. (Amended One Time) The method of claim 41, wherein classifying the fragment further comprises the steps of:

analyzing fragment for uniqueness; and

adding information to a database regarding the uniqueness of the fragment.

43. (Amended One Time) The method of claim 42, further comprising the step of:

classifying the fragment by detecting a duplicate fragment.

44. (Amended One Time) The method of claim 37, further comprising the step of:

interacting with a user interface that administers the system.

45. (Amended One Time) The method of claim 37, further comprising the step of:

including uniform resource locator information regarding said at least one Web site in the traffic data.

46. (Amended One Time) The method of claim 37, further comprising the step of:

performing data integrity monitoring of the sampling data.

47. (Amended One Time) The method of claim 37, further comprising the steps of:

performing automatic advertisement detection by applying at least one heuristic algorithm to detect advertising within an HTML or an XML document; and

normalizing the detected HTML or XML content into a hierarchical form.

48. (Amended One Time) A computer readable medium comprising:

code for computing a number of impressions of digital content in traffic data;

code for retrieving the digital content from the traffic data to generate sampling data; and

code for generating prevalence estimates for the digital content from the traffic data and the sampling data.

49. (Amended One Time) The computer readable medium of claim 48, further comprising:

code for retrieving a Web page from said at least one Web site;

code for extracting a fragment from the Web page; and

code to classify the fragment.

50. (Amended One Time) A system for estimating prevalence of digital content on a network, comprising:

means for obtaining traffic data from at least one Web site on the network;

means for computing a number of impressions for the digital content traffic data;

means for retrieving the digital content from the traffic data to generate sampling data; and

means for generating prevalence estimates of the digital content from the traffic data and the sampling data.

51. (Amended One Time) The system of claim 50, further comprising:

means for classifying a fragment extracted from a Web page.

#1 Concl.

52. (Amended One Time) The system of claim 50, further comprising:

means for anonymizing the traffic data.

53. (Amended One Time) A system of estimating prevalence of digital content on a network, comprising:

means for estimating global traffic to at least one Web site on the network to provide traffic data;

means for statistically sampling the contents of said at least one Web site to provide sampling data;

means for storing the traffic data and the sampling data; and

means for generating prevalence estimates for the digital content by accessing the traffic data and the sampling data.

New Claims

54. (New) The system of claim 53, further comprising:

means for reporting the prevalence estimates to a user.

55. (New) A method for using a computer to estimate prevalence of digital content on a network, comprising the steps of:

storing traffic data collected from the network;

storing summarized traffic data that describe each occurrence of the digital content in the traffic data; and

presenting the traffic data and the summarized traffic data to a user.

56. (New) The method of claim 55, wherein storing traffic data further comprises the step of:

retrieving the traffic data from at least one proxy server.

57. (New) The method of claim 55, wherein storing summarized traffic data further comprises the step of:

computing the number of impressions of the digital content for a web site on the network.

58. (New) The method of claim 55, wherein storing traffic data further comprises the steps of:

fetching a web page from the network;

locating a fragment of the web page that includes the digital content; and

performing a structural analysis of the fragment to classify the digital content.

59. (New) The method of claim 55, wherein presenting the traffic data and the summarized traffic data further comprises the step of:

generating a report when the traffic data or the summarized traffic data satisfy at least one criterion.

AY

60. (New) A system for estimating prevalence of digital content on a network, comprising:

a memory device; and

a processor disposed in communication with the memory device, the processor configured to:

store traffic data collected from the network;

store summarized traffic data that describe each occurrence of the digital content in the traffic data; and

present the traffic data and the summarized traffic data to a user.

61. (New) The system of claim 60, wherein the processor retrieves the traffic data from at least one proxy cache server.

62. (New) The system of claim 60, wherein the processor computes the number of impressions of the digital content for a web site on the network.

63. (New) The system of claim 60, wherein the processor is further configured to:

fetch a web page from the network;

locate a fragment of the web page that includes the digital content; and

perform a structural analysis of the fragment to classify the digital content.

64. (New) The system of claim 60, wherein the processor generates a report when the traffic data or the summarized traffic data satisfy at least one criterion.

65. (New) A computer readable medium comprising:

code for storing traffic data collected from the network;

code for storing summarized traffic data that describe each occurrence of the digital content in the traffic data; and

code for presenting the traffic data and the summarized traffic data to a user.

66. (New) The computer readable medium of claim 65, further comprising:

code for retrieving the traffic data from at least one proxy cache server.